AUTOMATED TREE MORTALITY DETECTION USING UBIQUITOUSLY AVAILABLE PUBLIC DATA

A Thesis

presented to

the Faculty of California Polytechnic State University

San Luis Obispo

In Partial Fulfillment

of the Requirements for the Degree

Master of Science in Biology

by

Michael Huggins

March 2024
COMMITTEE MEMBERSHIP

TITLE: Automated Tree Mortality Detection Using Ubiquitously Available Public Data

AUTHOR: Michael Huggins

DATE SUBMITTED: March 2024

COMMITTEE CHAIR: Jenn Yost, Ph.D.
Professor of Biology

COMMITTEE MEMBER: Matthew Ritter, Ph.D.
Professor of Biology

COMMITTEE MEMBER: G. Andrew Fricker, Ph.D.
Assistant Professor of Geography
ABSTRACT

Automated Tree Mortality Detection Using Ubiquitously Available Public Data

Michael Huggins

Understanding the dynamic interplay between fire severity, topography, and tree mortality, is crucial for predicting future forest dynamics and enhancing resilience against climate change-induced wildfire regimes. This thesis develops a multi-sensor approach for automated estimation of tree mortality, then applies it to examine trends in tree mortality over a six-year period across a fire affected study site in the Trinity River basin in Northern California. The Random Forest model uses publicly available USGS 3D Elevation Program Lidar (3DEP) and NAIP imagery as inputs and is likely to be easily adaptable to other landscapes. The model had a Receiver Operating Characteristic Area Under the Curve (ROC AUC) score in training of 0.998. In multiple rounds of validation, using geographically distinct sets of holdout data, had mean accuracy of 0.998. The trained model was then used to assess tree mortality across a patchwork of different levels of burn severity at a site in Northern California. When applied to the study site significant variations were found in tree mortality across different fire severity treatments and landforms. This model shows potential for incorporation into predictive tree mortality models based on landform and climate.
# TABLE OF CONTENTS

<table>
<thead>
<tr>
<th>LIST OF TABLES</th>
<th>vii</th>
</tr>
</thead>
<tbody>
<tr>
<td>LIST OF FIGURES</td>
<td>viii</td>
</tr>
<tr>
<td>CHAPTER</td>
<td>1:</td>
</tr>
</tbody>
</table>
| 1 INTRODUCTION                                    | 1:
| 1.1 Overview                                      | 1:  |
| 1.2 Previous Work                                 | 2:  |
| 1.3 Goals of Study                                | 5:  |
| 2 A MULTI-SENSOR RANDOM FOREST TREE MORTALITY CLASSIFICATION MODEL | 7:  |
| 2.1 Introduction                                  | 7:  |
| 2.2 Methods                                       | 8:  |
| 2.2.1 Sites                                       | 9:  |
| 2.2.2 Data                                        | 10: |
| 2.2.3 Tree Delineation                            | 12: |
| 2.2.4 Labeling of Data                            | 13: |
| 2.2.5 Feature Engineering                         | 13: |
| 2.2.6 Model Selection and Parameterization        | 15: |
| 2.2.7 Model Validation                            | 18: |
| 2.3 Results                                       | 19: |
| 2.3.1 Optimal Number of Features                  | 20: |
| 2.3.2 Amount of Training Data                     | 22: |
## LIST OF TABLES

<table>
<thead>
<tr>
<th>Table</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>2.1 PRISM 30 year normal (1991-2021) precipitation [14], with mean and standard deviation of elevation [43], for each tile used in the model training and validation process.</td>
<td>10</td>
</tr>
<tr>
<td>2.2 List of model features and their descriptions</td>
<td>17</td>
</tr>
<tr>
<td>2.3 Results using each tile and year combination as a validation holdout. Tile 10TDL0480045075 for the year 2018 does not contain any dead trees. ROC UAC score and Log Loss (as implemented in Scikit-Learn) cannot be computed with fewer than two classes represented in a data set.</td>
<td>23</td>
</tr>
<tr>
<td>3.1 Percentage of site landcover by SAF cover type. [38]</td>
<td>39</td>
</tr>
<tr>
<td>3.2 Number of crowns sampled from each treatment.</td>
<td>41</td>
</tr>
<tr>
<td>3.3 Maximum likelihood estimates for sample distributions by treatment and year. Distributions are shown graphically with 95% HDIs in figure 3.3</td>
<td>45</td>
</tr>
<tr>
<td>B.1 Results using each tile and year combination as a validation holdout. Tile 10TDL0480045075 for the year 2018 does not contain any dead trees. ROC UAC score and Log Loss (as implemented in Scikit-Learn) cannot be computed with fewer than two classes represented in a data set.</td>
<td>69</td>
</tr>
<tr>
<td>Figure</td>
<td>Description</td>
</tr>
<tr>
<td>--------</td>
<td>-------------</td>
</tr>
<tr>
<td>2.1</td>
<td>Left: Map showing locations of tiles used for model training and validation within the Trinity River Watershed (huc8: 18010211). Right: Map showing location of Trinity River Basin within surrounding counties and national forest.</td>
</tr>
<tr>
<td>2.2</td>
<td>RGI histograms for three example tree crowns. The value of each bin corresponds to a separate model input feature, e.g. the crown represented by the image on the left has a value of 0.74 for the feature \texttt{rgi40} and a value of 0.26 for the feature \texttt{rgi50}.</td>
</tr>
<tr>
<td>2.3</td>
<td>Confusion matrix for predictions on validation data. Darker colors indicate a higher count. The number of false positives and false negatives (18 each) were well balanced and formed a small proportion of the total. The validation accuracy is 0.985.</td>
</tr>
<tr>
<td>2.4</td>
<td>Relative importance of the top 50 parameters resulting from parameter tuning trial.</td>
</tr>
<tr>
<td>2.5</td>
<td>Shows the model performance, as ROC AUC, resulting from using different numbers of features with the model. For each marker shown, the $n$ most influential features, as determined by the parameter tuning trial, were used to train and validate the model. Model performance improved rapidly up until about 20 crowns, after which point the score appears to level out. The best score, of about 0.9988, occurs using 62 features (highlighted in red).</td>
</tr>
<tr>
<td>2.6</td>
<td>Shows the model performance, as ROC AUC, resulting from training the model on different sized data sets. Model performance improved rapidly as more crowns were added up until around 2,000 crowns, after which there is a slow improvement until about 9,000 crowns at which point the score appears to level out. The best score, occurs around 9,700 or so.</td>
</tr>
</tbody>
</table>
2.7 Scores of tiles (green dotted lines) are shown over the histogram (grey) of model prediction scores of random samples of crowns taken from the pooled data set of all labeled crowns. Pink vertical lines show the bounds of the 95% credible interval of the underlying distribution of scores from random samples. Only one tile (10TDL0488245360) falls outside of the 95% credible interval.

2.8 Kernel densities showing the distance of prediction probabilities from the classification threshold (0.5) grouped by correct and incorrect predictions. Correct predictions (blue line) tend to be very far from the threshold (mean distance = 0.489, where 5 is the maximum possible). Distances of the probabilities of incorrect predictions, on the other hand, are more disperse and, on average, much closer to the threshold (mean = 0.175). Note that negative values and those over 0.5 are artifacts of the kernel density function fitting. Actual values all lie between 0 and 0.5.

2.9 The frequency of occurrence of parameter values for the model tuning runs (n=11). Parameter values are shown on the y-axis. Variation along the x-axis is jitter introduced to make the individual data points distinct. The number of estimators was allowed to range from 2–128. Values were clustered at the high end of this range (rf_n_estimators ranged from 87–128, with a median value of 114). This suggests that if the range had been set higher some models may have ended up with more estimators.

2.10 Hyperparameter importances for the first (top) and last (bottom) tuning trials. In the first trial the number of estimators was relatively more important than in the last (0.66 and 0.52) and correspondingly the maximum depth of trees was less important in the first than in the last (0.28 and 0.43). The maximum number of features used per tree was relatively unimportant in both runs (0.06 and 0.05). Relative importances for other trials were not recorded.

2.11 Correlation between the 62 features used in the model. Features were sorted by importance, with correlation between the most important features shown in the upper left corner of the triangle.
2.12 Contour plots of ROC AUC resulting from different combinations of \texttt{rf_n_estimators} and \texttt{rf_max_depth} for the first (top) and last (bottom) tuning trials. The y-axis corresponds to the number of estimators, and the x-axis to the maximum permitted depth of estimators. ROC-AUC is shown by color, with deeper blue corresponding to higher scores. The pattern of points shows combinations of the two values visited during the tuning process. It can be seen that in both tuning runs, there were many models with extremely high ROC-AUC scores.

3.1 Shows site location and treatments. 
   a) Close up of study sites showing areas of the four treatments under consideration.
   b) Location of site within California.
   c) Location of site within Trinity County. Note that the Trinity River flows through the southeastern corner of the study site (panel c). Most of the area burned by the Monument fire, but not the Helena fire lies to the south-west of the river.

3.2 Top: Boxplot summarizing the distributions of the sample population for the entire site based on the RF model as described in section 3.3. The percentage of crowns classified as dead varies substantially by year. Bottom: Annual rainfall [34] is shown by the blue line. PRISM 30 year normal precipitation [14] for the same point is shown by the dotted red line. The years shown in the upper panel are marked by the vertical red lines.

3.3 Shows mortality for each year and treatment under consideration. All differences by year are significant, with the exception of 2018 and 2022 in the unburned — unburned treatment. Maximum likelihood estimates for the distributions shown here can be found in table 3.3.

3.4 Shows number of significantly different groups by year and geomorphons scale. The number of groups tends to increase with scale (and, interestingly, also as time progresses). Scale of 2000 consistently has the most separation of groups.
3.5 Relative composition, by landform, of crowns in each treatment group. Amongst the unburned groups (top row) unburned-unburned has a higher percentage of the more mesic landforms (hollows, valleys, and depressions) while unburned-high has higher percentage of the more xeric landforms (summit, ridge, spur, slope) across the three scales shown (the exception to this trend is slope, at scale = 100). Amongst the burned groups (bottom row of figure 3.5) The trend is less distinct, but in general the high-unburned treatment group has a higher proportion of the more mesic landforms and high-high have a higher proportion of the more xeric landforms. Ridges and valleys with scale = 2000 do not conform to this trend. The high-high treatment does not contain any crowns found on summit or depression landforms...
B.7 Confusion matrix using tile 10TDL0480045075 for 2018 as holdout data. Darker colors indicate a higher count.  

B.8 Confusion matrix using tile 10TEK0500244992 for 2022 as holdout data. Darker colors indicate a higher count.  

B.9 Confusion matrix using tile 10TDL0488245360 for 2022 as holdout data. Darker colors indicate a higher count.  

C.1 Percent of crowns dead in simulated populations for 2018 for each landform (using geomorphons with outer radius of 100 and inner radius of 50). 95% HDIs are shown by black bars with labels. Depressions and valleys were significantly different than all other landforms. All other landforms overlapped.  

C.2 Percent of crowns dead in simulated populations for 2018 for each landform (using geomorphons with outer radius of 250 and inner radius of 125). 95% HDIs are shown by black bars with labels. Depressions and valleys were significantly different than all other landforms. All other landforms overlapped some others.  

C.3 Percent of crowns dead in simulated populations for 2018 for each landform (using geomorphons with outer radius of 500 and inner radius of 250). 95% HDIs are shown by black bars with labels. Depressions and valleys were significantly different than all other landforms. Other landform overlapped to various degrees.  

C.4 Percent of crowns dead in simulated populations for 2018 for each landform (using geomorphons with outer radius of 1000 and inner radius of 500). 95% HDIs are shown by black bars with labels. Depressions and valleys were significantly different than all other landforms. Spurs, slopes and hollows formed a cluster, as did summits and ridges.  

C.5 Percent of crowns dead in simulated populations for 2018 for each landform (using geomorphons with outer radius of 2000 and inner radius of 1000). 95% HDIs are shown by black bars with labels. Depressions, valleys, and ridges were significantly different from all other landforms. Summits, spurs, slopes, and hollows formed a cluster.  

C.6 Percent of crowns dead in simulated populations for 2020 for each landform (using geomorphons with outer radius of 100 and inner radius of 50). 95% HDIs are shown by black bars with labels. Depressions and valleys were significantly different than all other landforms. Other landforms overlapped one another.
C.7 Percent of crowns dead in simulated populations for 2020 for each landform (using geomorphons with outer radius of 250 and inner radius of 125). 95% HDIs are shown by black bars with labels. Depressions and valleys were significantly different than all other landforms. Other landforms overlapped one another to varying degrees.

C.8 Percent of crowns dead in simulated populations for 2020 for each landform (using geomorphons with outer radius of 500 and inner radius of 250). 95% HDIs are shown by black bars with labels. Depressions, valleys, and slopes were significantly different than all other landforms. Other landforms overlapped one another to varying degrees, though summits were almost significant.

C.9 Percent of crowns dead in simulated populations for 2020 for each landform (using geomorphons with outer radius of 1000 and inner radius of 500). 95% HDIs are shown by black bars with labels. Depressions and valleys were significantly different than all other landforms. Other landforms overlapped one another to varying degrees.

C.10 Percent of crowns dead in simulated populations for 2020 for each landform (using geomorphons with outer radius of 2000 and inner radius of 1000). 95% HDIs are shown by black bars with labels. Depressions, valleys, and ridges were significantly different from all other landforms. Summits, spurs, slopes, and hollows formed a cluster.

C.11 Percent of crowns dead in simulated populations for 2022 for each landform (using geomorphons with outer radius of 100 and inner radius of 50). 95% HDIs are shown by black bars with labels. Depressions and valleys were significantly different than all other landforms. Other landforms overlapped one another to varying degrees.

C.12 Percent of crowns dead in simulated populations for 2022 for each landform (using geomorphons with outer radius of 250 and inner radius of 125). 95% HDIs are shown by black bars with labels. Depressions and valleys were significantly different than all other landforms. Other landforms overlapped one another to varying degrees.

C.13 Percent of crowns dead in simulated populations for 2022 for each landform (using geomorphons with outer radius of 500 and inner radius of 250). 95% HDIs are shown by black bars with labels. Depressions, valleys, and hollows were significantly different than all other landforms. Other landforms overlapped one another to varying degrees.
C.14 Percent of crowns dead in simulated populations for 2022 for each landform (using geomorphons with outer radius of 1000 and inner radius of 500). 95% HDIs are shown by black bars with labels. Depressions, valleys and spurs were significantly different than all other landforms. Other landforms overlapped one another to varying degrees.

C.15 Percent of crowns dead in simulated populations for 2022 for each landform (using geomorphons with outer radius of 2000 and inner radius of 1000). 95% HDIs are shown by black bars with labels. Depressions, valleys, hollows and slopes were significantly different than all other landforms. Summits, ridges, and spurs formed a mutually overlapping cluster.
CHAPTER 1

INTRODUCTION

1.1 Overview

Forests and woodlands play pivotal roles in the Earth’s ecological balance, acting as vital carbon sinks that mitigate climate change by absorbing significant amounts of CO$_2$ from the atmosphere. These complex ecosystems are critical natural resources, providing essential services including biodiversity conservation, water regulation, and soil stabilization. Unfortunately, many forests and woodlands are under increasing stress due to climate change, anthropogenic land use change, and human attempts at forest management. Understanding tree mortality through advanced techniques like remote sensing is crucial for forest management and conservation strategies. Such understanding is needed for planning of adaptation in order to mitigate environmental stresses and disturbances and maintain the health and resilience of forest ecosystems, which are indispensable for ecological stability, human well-being, and the fight against global climate change.

Remote sensing has been applied to forest monitoring for decades at this point, evolving from simple aerial photographs to sophisticated multispectral and hyperspectral imaging. Many reliable methods have been developed in the realm of remote sensing using medium resolution imagery such as Landsat and Sentinel [46]. Examples include the use of the tasseled cap transformation with Landsat or Sentinel data, which helps in distinguishing vegetation types, assessing their condition, and monitoring changes over time [40, 17] along with various methods of segmentation for use...
with object-based image analysis (OBIA) [33]. Another common method is to use higher resolution imagery in tree mortality surveys and then use the results to train models using medium resolution imagery. For example Hart et al. (2015) [15] used National Agriculture Imagery Program (NAIP) imagery for this purpose.

Unlike traditional spectral imaging, Light Detection and Ranging (LiDAR) provides detailed three-dimensional information by measuring the time it takes for pulsed laser light to return after reflecting off surfaces. LiDAR offers useful detail in depicting forest structure, including canopy height, density, and the vertical distribution of foliage.

LiDAR’s ability to penetrate the forest canopy allows for the accurate measurement of understory and ground layers, offering insights into forest biomass and ground topography that are often obscured in spectral imagery. The high-resolution and spatial accuracy of LiDAR data improves detection of individual trees and the mapping of gaps within the canopy. It is possible to calculate many other cloud or (after segmentation) object based metrics from point clouds such as the cumulative percentage of returns in a layer where layers are determined by dividing the range of height measurements into equal intervals [45], percentiles of the height distribution, entropy, and others [36].

1.2 Previous Work

Several researches have used LiDAR in combination with other forms of remote sensing to assess tree mortality. Cambell et al. (2020) [5] employed a multi-sensor approach to analyze tree mortality in piñon-juniper woodlands. Initially, they utilized LiDAR data to create a detailed three-dimensional representation of the woodland structure, enabling the delineation of individual tree crowns. Then they used high-resolution imagery from an Unmanned Aircraft System (UAS) to visually assess and
categorize individual trees as alive or dead. They then used Landsat 8 OLI imagery derived vegetation indices from before and after a mortality event for the detection of vegetation cover changes over the broader landscape. The study integrated these datasets using a stepwise regression model to link localized, high-resolution mortality assessments to satellite-derived vegetation changes. Impressively, the model was able to predict Pinion and Juniper mortality within super pixels quite well with a mean absolute error (MAE) of 6.92%. The model employs a novel, tree segmentation algorithm, which seems to be primarily suited to savannah or open woodlands, and perhaps even more specifically piñon-juniper woodlands. It also, as mentioned required the collection of UAV imagery, which, though easy in many cases is not always a possibility.

In their study at Cal Poly’s Swanton Pacific Ranch, Bishop et al. (2014) [4] assessed the utility of combining NAIP imagery and LiDAR data to estimate tree mortality following a wildfire. The analysis was based on 47 continuous forest inventory plots, with mortality classifications derived from different combinations NAIP imagery and LiDAR-derived cloud metrics. They found that using post-fire LiDAR data in combination with NAIP imagery achieved an overall accuracy of 85% in classifying plot-level tree mortality. This model significantly outperformed the imagery-only approach, which had a 74% accuracy. Interestingly the addition of pre-fire LiDAR data to examine differences in LiDAR before and after the fire showed a slight decrease in performance compared to using only post-fire LiDAR data, with an overall accuracy of 83%.

Hemming-Schroeder et al. (2023) [18] examined tree mortality at the National Ecological Observatory Network (NEON) Soaproot Saddle and Lower Teakettle sites in the Sierra Nevada of California. They made use of the high-resolution LiDAR and hyperspectral data from NEON’s airborne observation platform. The research utilized comprehensive datasets spanning 2013 to 2021, including LiDAR point clouds
to create canopy height models and derive topographical features, as well as hyperspectral reflectance data to generate vegetation indices and classify tree mortality. Hyperspectral data was used to create vegetation indices, such as relative greenness and NDVI, to identify live and dead trees. A significant part of the analysis was the development of algorithms for tree crown delineation and classification based on combined LiDAR datasets, spanning several years, enabling the tracking of individual trees over time. The study also introduces a granite-detection algorithm to explore the impact of surrounding granite on tree health. The mortality classifier had an accuracy of 93.3%. It is worth noting that the training and validation data was manually labeled using RGB imagery derived from NEON’s hyperspectral data. This manual labeling process was based on specific criteria related to tree crown area and coloration changes over time. As such the classifier was being tested on its ability to emulate human hand labeled training data rather than actual ground-truthed mortality. As such, this study is the most similar example to what is presented in this study.

Fricker et al. (2019) [13] also combine NEON LiDAR and hyperspectral imagery at the Teakettle experimental forest. Similarly to this study, they use LiDAR to locate trees. They use a convolution neural network and a pixel based approach for classification of tree species, as well as alive/dead status. They found that for species identification the hyperspectral imagery dramatically outperformed RGB imagery.

Two more methods worth mention that are commonly used in tree mortality surveys, and will be discussed in more detail in the introduction to chapter 2, are sketch mapping, in which analysts look out the windows of airplanes and mark observed areas of damage using purpose built, digital aerial sketch-mapping tablet based GIS systems, and photo interpretation which also involves analysts, but working on a desktop workstation using arial imagery rather than in real time as a plane is flying over the area of interest.
1.3 Goals of Study

The studies above show the utility of multi-sensor approaches combining LiDAR and spectral data for monitoring forest health. However the quality of data in these studies is not available in many cases. In the above examples data availability ranges from ideal [18] to very good [4]. Rarely is co-registered, annual, LiDAR and hyperspectral data available for a site (which is not a NEON observatory) as was the case for Hemming-Schroeder et al. (2023) [18]. Bishop et al. (2014) [4] had two LiDAR collections two years apart, with a major disturbance in between. Cambell et al. (2020) [5] collected high resolution UAV data to use in their classification algorithm.

In this thesis I present a model for individual tree level mortality detection using ubiquitous, publicly available data. This model should be easily adaptable to most parts of the continental United States (CONUS). The site used here is likely more challenging than most due to its steep topography and dense tree cover in some areas. The goal of the model is to automatically classify tree crowns as living or dead, based solely on USGS 3D Elevation Program (3DEP) LiDAR, and NAIP imagery.

The 3DEP and NAIP datasets were chosen because they are free and publicly available. NAIP has high enough resolution to discern individual crowns (60 cm in more recent years, 1 m in older data) and is available in most areas of the continental United States every other year. The USGS 3DEP LiDAR collections are the only widely available LiDAR datasets in most parts of the United States. They come in several quality levels (QL0, QL1, etc...) depending on the exact purpose under which they were collected. The USGS has the stated goal of releasing 3DEP data on no more than an eight year interval for any given location in the continental United States.

As of writing, the study area, The Trinity River Basin (in Klamath Mountains of Northern California), has only one recent, wide-coverage, LiDAR collection (the
USGS CarrHirzDeltaFire 2019 collection [41], though a 2022 collection will be forthcoming).

In Chapter 2 I introduce the model, discuss the details of the algorithm, the training and the validation of the model. I hypothesize that 1) by taking advantage of the forest structural information present in the LiDAR data, combining it with the 60 cm resolution, red, green, blue, infrared (RGBI) NAIP imagery, and using machine learning techniques, it will be possible to classify individual tree crowns as living or dead. 2) The accuracy of the model’s classification of living or dead trees will be close to the accuracy of hand labeled training datasets used to train the model.

In chapter 3 I demonstrate a simplistic application of the model wherein I ask the questions: 1) Using the model, is it possible to discern differences in tree mortality based on recent fire history?; 2) Using the model, is it possible to discern differences in tree mortality based on topographic position?; 3) What is the interaction between topographic position and recent fire history in determining mortality rates?

SectionCode Availability Important elements of the Python code used in both chapters 2 and 3 can be found at https://github.com/kulpojke/tree-mortality/blob/main/html_notebooks. An example (the same code was reused with different file names many times) of R code used in tree segmentation can be found at https://github.com/kulpojke/tree-mortality/blob/main/src/helena_tiles.R
CHAPTER 2

A MULTI-SENSOR RANDOM FOREST TREE MORTALITY CLASSIFICATION MODEL

2.1 Introduction

Currently tree mortality or damage surveys, such as the United States Forest Service Aerial Detection Survey (ADS) [42], gather data by a technique known as sketchmapping, in which analysts look out the windows of airplanes and mark observed areas of damage using purpose built, digital aerial sketch-mapping tablet based GIS system. ADS results in a map with coarse polygons which contain areas of tree damage, and estimates for how many trees per area are dead or damaged.

Photo interpretation surveys can be performed using similar techniques to those employed in ADS, but on a desktop workstation using aerial imagery. While the imagery may not allow for seeing as much detail, it removes the time pressure for the technician, as they can examine the photo for a long as they need to. Backsen et al. [3] found that, for their study site in the Black Hills of South Dakota, both ADS and photo interpretation surveys underestimated (by $\sim 82\%$ and $\sim 76\%$ respectively) trees per area killed as compared to surveys of ground plots.

In this chapter I present a model for quantifying individual tree mortality—for use in producing area based estimates of tree mortality—using USGS 3D Elevation Program (3DEP) LiDAR [41] and National Agricultural Imagery Program (NAIP, [31]) four band spectral data. I hypothesize that 1) by taking advantage of the forest structural information present in the LiDAR data, combining it with the 60
cm resolution, red, green, blue, infrared (RGBI) NAIP imagery, and using machine learning techniques, it will be possible to classify individual tree crowns as living or dead. 2) The accuracy of the model’s classification of living or dead trees will be close to the accuracy of hand labeled (e.g. phot interpreted) training data sets used to train the model.

The model presented here is designed for use in Trinity River Basin, and more generally Klamath Mountains, however the framework could be applied in other contexts if trained on the appropriate data. It relies on the LiDAR data to find tree crowns, then classifies them using spectral information from the NAIP imagery. Currently the only 3DEP collection available for the site is the 2019 CarrHirzDeltaFire data set [41], however the site will be covered by the 2022 Northern California 3DEP collection, which will most likely be available sometime in 2024. When the 2022 LiDAR is available it will be possible to augment the mortality classification model using the changes in tree crowns found in the two LiDAR data sets (as done, for example by Hemming-Schroeder et al. (2023) [18] and Bishop et al. (2014) [4]).

2.2 Methods

The final version of the model presented in this chapter is a Random Forest (RF) model (though Support Vector Classifiers (SVC) were also considered). The model uses LiDAR data to find tree crowns, then classifies them using spectral information from the NAIP imagery. It is designed to be agnostic to the exact resolution of the imagery. This was done so that it can be used with newer or older NAIP imagery, or if desired imagery collected by UAV or other platforms. The model does rely on the imagery being well registered to the LiDAR data. It works out of the box with NAIP (at least in this case), but for other imagery sources, such as Planet, the imagery would likely need to be co-registered with the LiDAR. It is also important that the
model be able to make robust predictions in images with varying degrees of shadow and/or bare soil, and work on tree crowns of various sizes without pre-processing the images. In the study area bare soil is often present around and within the tree crowns, shadows are ubiquitous and at varying angles, and tree crowns are often only partially within their LiDAR based polygon representation due to parallax.

2.2.1 Sites

The sites used for training and validation of the model are situated within the Klamath Mountains. Each site is contained within a single tile of the USGS CarrHirzDeltaFire LiDAR data set [41]. With the exception of tile 10TEK0503244655—which is in the Cottonwood Creek Basin (Eight digit hydrologic unit code (huc8): 18020152), about 8 km south of the southern tip of the Trinity River Basin—all of the tiles used in training and validation are located in the Trinity River basin (huc8: 18010211). The tiles were manually selected based on the presence of both living and dead trees in 2022 NAIP imagery.

The sites range in elevation from 489 to 1946 m [43]. The 30 year normal precipitation from the Parameter-elevated Regressions on Independent Slopes Model (PRISM) for the period 1991-2021, ranges from 898 to 1585 mm y$^{-1}$ [14] (see table 2.1). Six of the sites are within Trinity County, the other, 10TEK0503244655, is in south-western Shasta County. Four of the sites are situated entirely within the Shasta-Trinity National forest and one, 10TDL0458245240, is completely within the Six Rivers National Forest. Tile 10TEK0503244655 is mostly in the Shasta-Trinity National Forest with the eastern portion extending onto BLM lands. The seventh site, 10TEK0500244992, is located on private land—most of which is designated as timber production—in eastern Trinity County. Site locations can be seen in Figure 2.1.
Vegetation in the region is dominated by conifer forests and associated broad-leaved evergreen trees [10]. Society of American Foresters (SAF) land cover classes (taken from USFS E\text{Veg} Mid Region 5 North Coast East data set [38] which is based on input imagery requisitioned between 1998 and 2015) for the tiles used in the training data can be seen in appendix A. The training tiles contain the following SAF land-cover types; Not forest or woodland, California black oak, canyon live oak, Douglas-fir - tanoak - Pacific madrone, hard chaparral, Oregon white oak, Pacific Douglas-fir, Pacific Douglas-fir, Pacific ponderosa pine, Pacific ponderosa pine - Douglas-fir, red fir, Sierra Nevada mixed conifer, white fir.

Table 2.1 – PRISM 30 year normal (1991 -2021) precipitation [14], with mean and standard deviation of elevation [43], for each tile used in the model training and validation process.

<table>
<thead>
<tr>
<th>Tile</th>
<th>30 Yr Normal Precip. (mm)</th>
<th>Mean Elev. (m)</th>
<th>Elev. stdev (m)</th>
</tr>
</thead>
<tbody>
<tr>
<td>10TEL0509245547</td>
<td>1585</td>
<td>1946</td>
<td>74</td>
</tr>
<tr>
<td>10TEK0500244992</td>
<td>898</td>
<td>699</td>
<td>79</td>
</tr>
<tr>
<td>10TEK0503244655</td>
<td>950</td>
<td>945</td>
<td>44</td>
</tr>
<tr>
<td>10TDL0480045075</td>
<td>922</td>
<td>790</td>
<td>47</td>
</tr>
<tr>
<td>10TDL0488245360</td>
<td>1157</td>
<td>1241</td>
<td>71</td>
</tr>
<tr>
<td>10TDL0458245240</td>
<td>1296</td>
<td>489</td>
<td>94</td>
</tr>
<tr>
<td>10TDL0464245187</td>
<td>1190</td>
<td>1044</td>
<td>66</td>
</tr>
</tbody>
</table>

2.2.2 Data

The 3DEP and NAIP data sets were chosen because they are free and publicly available. NAIP has high enough resolution to discern individual crowns (60 cm in more recent years, 1 m in older data) and is available in most areas of the continental United States every other year. The USGS 3DEP LiDAR collections are the only widely available LiDAR data sets in most parts of the United States. They come in several quality levels (QL0, QL1, etc...) depending on the exact purpose under...
which they were collected. The USGS has the stated goal of releasing 3DEP data on no more than an eight year interval for any given location in the continental United States.

In particular, the LiDAR data used in the study comes from the 2019 USGS CarrHirzDeltaFire collection [41]. NAIP four-band imagery [31] for the years 2018, 2020, and 2022, was downloaded using the USGS EROS explorer (Accessed 2023-07-28).
2.2.3 Tree Delineation

Approximations of tree crowns were delineated using the CarrHirzDeltaFire LiDAR data set [41] with the lidR package [36] in the R statistical programming language [35]. The specific method employed is the Dalponte algorithm [11], which uses region growing techniques on a Canopy Height Model (CHM). Several steps of pre-processing the LiDAR data were undertaken within the lidR package to create and smooth the CHM before the actual segmentation.

The x, y, and z coordinates as well as intensity, classification, and return number were read from the laz file. Points flagged as withheld were ignored, and duplicate points were filtered at read-time. The point cloud was then normalized using the `normalize_height` function. Points with normalized height below 3 m and above 120 were then filtered out in order to reduce the number of returns not within tree crowns. A 0.5 m resolution CHM was made from the normalized and filtered pointcloud using the `rasterize_canopy` function. The CHM was then smoothed using a 3×3 pixel window mean filter. Tree tops were found from the smoothed CHM using `locate_trees` with a variable window defined by equation (2.1) after Hemming et al.[18].

\[
  w(h) = \begin{cases} 
    4 \text{ m}, & \text{if } h \leq 32 \text{ m} \\
    \frac{h}{8}, & \text{if } 32 < h < 80 \text{ m} \\
    10 \text{ m}, & \text{if } h \geq 80 \text{ m}
  \end{cases} \tag{2.1}
\]

where \( w \) is the window size, and \( h \) is the height, both in m.

Next, the normalized point cloud was segmented into individual tree crowns—such that each point belongs to exactly one tree—using the tree tops and smoothed CHM as inputs to the `segment_trees` function with the `dalponte2016` algorithm [11]. Metrics were calculated for each crown using `crown_metrics` with the `.stdmetrics`
function. Tree tops and crowns, with their associated metrics were then saved as geopackages.

2.2.4 Labeling of Data

Since ground-truth data was unavailable for the area, training data was created by visually inspecting imagery and manually labeling the crown polygons based on subjective belief as to whether a tree is alive or dead. Originally I planned on using a scoring rubric similar to that used by Hemming et al. (2023) [18], however this proved impractical due to the high percentage of pixels falling into shadow. Ways in which the labeling process could be be improved will be discussed in Section 2.4.

2.2.5 Feature Engineering

The spectral data available for each tree crown from NAIP imagery is a collection of pixels with four band values, red, green, blue and near infrared. The number of pixels within each crown is variable, depending on the size of the tree, and there are differences from year to year and scene to scene due to shadows, camera angle, and atmospheric conditions [12].

Another challenge, which will be examined more thoroughly in Section 2.4, is registration error of tree polygons and NAIP imagery. The registration error results in an artificial variation in bare-soil or neighboring tree pixels included within the polygon.

In order to facilitate classification with either a random forest or support vector machine based model it was necessary to engineer new variables from the data that could be used by the models. These features needed to capture the spectral data in a form where it relatively uncontaminated by shadows and bare soil. The features also need to minimize year to year and scene to seen differences cause by lighting and
atmospheric conditions. Furthermore, while minimizing the impact of bare soil and shadow on the spectral information, I did not want to completely discard information about the quantity of shadow and bare soil present within crown polygons. (This is in contrast to previous studies such as Bishop et al. (2014) [4] in which shadow pixels are adjusted by comparison with neighboring pixels.) With these goals in mind, I created a large vector of statistics describing the pixel population for a given crown. The steps used to create the feature vector of statistics is described below. Descriptions of each individual statistic can be found in table 2.2. A feature vector was created from the NAIP data, for each crown, for each year under consideration.

Upon opening the NAIP image for a given year, the 4 bands were normalized to the scale of the maximum value of any band in any pixel. Then several indices were calculated. Relative greenness (RGI) was calculated as,

\[ \text{RGI} = \frac{\text{green}}{\text{red} + \text{green} + \text{blue}} \]  \hspace{1cm} (2.2)

The \text{ndvi} function from Xarray-Spatial [27] was used to calculate NDVI as,

\[ \text{NDVI} = \frac{\text{NIR} - \text{red}}{\text{NIR} + \text{red}} \]  \hspace{1cm} (2.3)

and soil adjusted vegetation index (SAVI) [19] was calculated using the \text{savi} function from Xarray-Spatial, as,

\[ \text{SAVI} = \frac{\text{NIR} - \text{red}}{\text{NIR} + \text{red} + L} \times (1 + L) \]  \hspace{1cm} (2.4)

where \( L \), is a correction factor for reducing the influence of bare soil. In this case, \( L \) was set to one. This was done for two reasons, the first being that much of the soil in the area is highly reflective, the second being to maximize the difference between the values of SAVI and those of NDVI. Additionally, for each band a pixel normalized band was created (\text{red}\_norm, \text{green}\_norm, \text{blue}\_norm, \text{NIR}\_norm) for each color as,

\[ \text{band}\_\text{norm} = \frac{\text{band}}{\text{red} + \text{green} + \text{blue} + \text{NIR}} \]  \hspace{1cm} (2.5)
where \textit{band} is the band in question.

A mask to remove shadows and bare soil was created by calculating RGB intensity,

\[
intensity = \frac{\text{red} + \text{green} + \text{blue}}{3} \quad (2.6)
\]

then masking pixels with intensity lower than 0.176 as shadow, and those with intensity higher than 0.569 as bare soil. The threshold values for shadow and soil were acquired through trial and error masking and visual inspection of images.

Intensity, RGI, and each of the normalized bands have values ranging from zero to one. A histogram—divided into ten bins \((0 \leq n < 0.1, 0.1 \leq n < 0.2, \ldots , 0.9 \leq n < 1)\)—of pixel values was created for each of these quantities. The number of pixels falling within each bin was then used as a feature. As an example Figure 2.2 shows histograms of RGI for three crowns. Each column in the histogram corresponds to a model input feature.

The means and standard deviations were calculated for each of the indices and normalized bands and were included as features as well. Table 2.2 summarizes all of the model input features.

\subsection*{2.2.6 Model Selection and Parameterization}

All of the labeled crowns were read and pooled into a single dataframe using the GeoPandas package [22]. Features were created as described in Section 2.2.5. The resulting data set was then split into training and test data sets using an 80/20 split. The training set was then used for model selection.

Random forest classifiers (RF) and C-support vector classifiers (SVC) [6] were considered as possible models. The RF and SVC models were chosen for ease of implementation and their moderate computation expense. In early phases of exploring models, RF models performed extremely well, which served as a disincentive to further exploration of other models.
Figure 2.2 – RGI histograms for three example tree crowns. The value of each bin corresponds to a separate model input feature, e.g. the crown represented by the image on the left has a value of 0.74 for the feature \texttt{rgi40} and a value of 0.26 for the feature \texttt{rgi50}.

The Optuna library [1] was used to run 10,000 trials in order to optimize parameters of the RF and SVC models. Within each trial three-fold cross validation was used by optuna to select the best successful version of candidate models. Both models—\texttt{sklearn.ensemble.RandomForestClassifier} and \texttt{sklearn.svm.SVC} respectively—are from the scikit-learn library [32]. SVC models were trialed using \texttt{rbf} and \texttt{sigmoid} kernels, with the regularization parameter, \texttt{C}, ranging from $10^{-10}$ to $10^{10}$. The kernel coefficient, \texttt{gamma}, was set to \texttt{auto}, and the \texttt{class_weights} parameter was provide using a dictionary of the actual class weights.

For the RF model, values from the maximum depth of the trees, \texttt{max_depth}, were drawn from the interval [2,32] using Optuna with log sampling in order to favor lower values, while the number of trees in the forest, \texttt{n_estimators}, was examined over
Table 2.2 – List of model features and their descriptions

<table>
<thead>
<tr>
<th>Feature</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>lum10, lum20, ..., lum100</td>
<td>Fraction of pixels having intensity between successive intervals of 0.1 from 0 to 1.</td>
</tr>
<tr>
<td>rgi10, rgi20, ..., rgi100</td>
<td>Fraction of pixels having relative greenness between successive intervals of 0.1 from 0 to 1.</td>
</tr>
<tr>
<td>r10, r20, ..., r100</td>
<td>Fraction of pixels having red, normalized at pixel level, between successive intervals of 0.1 from 0 to 1.</td>
</tr>
<tr>
<td>g10, g20, ..., g100</td>
<td>Fraction of pixels having green, normalized at pixel level, between successive intervals of 0.1 from 0 to 1.</td>
</tr>
<tr>
<td>b10, b20, ..., b100</td>
<td>Fraction of pixels having blue, normalized at pixel level, between successive intervals of 0.1 from 0 to 1.</td>
</tr>
<tr>
<td>n10, n20, ..., n100</td>
<td>Fraction of pixels having NIR, normalized at pixel level, between successive intervals of 0.1 from 0 to 1.</td>
</tr>
<tr>
<td>ndvi_mean, ndvi_std</td>
<td>Mean and standard deviation of NDVI of all masked pixels within the crown.</td>
</tr>
<tr>
<td>rgi_mean, rgi_std</td>
<td>Mean and standard deviation of relative greenness of all masked pixels within the crown.</td>
</tr>
<tr>
<td>savi_mean, savi_std</td>
<td>Mean and standard deviation of SAVI of all pixels within the crown (SAVI used unmasked data).</td>
</tr>
<tr>
<td>r_mean, r_std</td>
<td>Mean and standard deviation of red, normalized at pixel level, of all masked pixels within the crown.</td>
</tr>
<tr>
<td>g_mean, g_std</td>
<td>Mean and standard deviation of green, normalized at pixel level, of all masked pixels within the crown.</td>
</tr>
<tr>
<td>b_mean, b_std</td>
<td>Mean and standard deviation of blue, normalized at pixel level, of all masked pixels within the crown.</td>
</tr>
<tr>
<td>n_mean, n_std</td>
<td>Mean and standard deviation of NIR, normalized at pixel level, of all masked pixels within the crown.</td>
</tr>
</tbody>
</table>

the range of [2,128] without log sampling. The rationale for using the chosen ranges of depth and number of trees is that a forest with more trees and shallower depth is less likely to be over-fit [30]. The number of features to consider when looking for the best split, max_features, ranged from 1, to 74 (all of the features) using log sampling to favor smaller values for performance reasons. In order to make results more reproducible, the seed for the pseudorandom number generator, random_state, (which controls the sampling of the features to consider when looking for the best split at each node) was specified (as 1234).
2.2.7 Model Validation

After the best model was selected the importance of its parameters were examined. Figure 2.4 shows the importance of the 50 most important parameters within all members of the ensemble. Importance drops off sharply for the first 13 or so parameters, then slowly declines until it is negligible. As such it seemed likely to me that the number of features used in the model could be reduced without loss of predictive power. In order to determine if this was true, the model was trained with different numbers of features and then scored using ROC AUC based on the test data. For each of these trials the first $n$ features, where $n$ ranged from 10 to 74 (i.e. all features), ranked by feature importance, were included in the model. As can be seen in Figure 2.5, the highest score was attained by using 62 features. While a more rigorous test could be devised for determining the number of features to use, it would likely not make much difference in overall model performance since the differences in ROC AUC score between different models were a few thousandths.

An argument could be made for using fewer than ten features in the model. In retrospect it would have made sense to begin at $n=2$ and hopefully in the future this should be examined. In the case of an explanatory model, a more parsimonious model is obviously preferable, however using more parameters will allow for a more diverse set of classifiers within the RF. As discussed in 2.2.6 and previously by Nadi et al. (2019) [30] a forest with more trees and shallower depth is less likely to be over-fit. Intuitively it seems that a more diverse set of trees may also help guard against overfitting, however a thorough investigation of this idea is beyond the scope of this paper.

Next, in order to look for biases introduced by differences in crown populations based on location or other tile-specific traits, the trained model was run on a tile by tile basis. The results (given in table 2.3) were scored using ROC AUC, accuracy,
and log-loss. Log-loss was used because I wanted to evaluate how certain the model is in its correct and incorrect class probabilities. Log-loss compares the prediction probabilities to the labels and penalizes them based on distance. For example, a true positive prediction which is more certain (e.g. a prediction probability of 0.9) is penalized less than a true positive prediction that is less certain (e.g. 0.6). A log-loss of zero is a perfect score.

2.3 Results

![Confusion Matrix for Test Data](image)

Figure 2.3 – Confusion matrix for predictions on validation data. Darker colors indicate a higher count. The number of false positives and false negatives (18 each) were well balanced and formed a small proportion of the total. The validation accuracy is 0.985.

The best model returned from the trial was a random forest classifier with 111 trees, maximum depth of 16, and using a maximum of 9 features per tree. It had a
ROC AUC score of 0.9979 from the three-fold cross-validation used by Optuna in the trial.

2.3.1 Optimal Number of Features

![Importance of Top 50 Parameters](image)

**Figure 2.4** – Relative importance of the top 50 parameters resulting from parameter tuning trial.

The best model was then used to make predictions on the test data, yielding ROC AUC validation score of 0.9982 and validation accuracy of 0.9858. The confusion matrix for predictions made on the validation data set is shown in Figure 2.3.

Figure 2.4 shows the importance of the 50 most important parameters within all members of the ensemble. Parameter importance drops off sharply for the first 13 or so parameters, then slowly declines until it is negligible. As such it seemed likely to me that the number of features used in the model could be reduced without loss of predictive power. In order to determine if this was true, the model was trained with different numbers of features and then scored using ROC AUC based on the test data. For each of these trials the first $n$ features (evaluated from 10 to 74, i.e.
all features) ranked by feature importance, were included in the model. As can be seen in Figure 2.5, the highest score was attained by using 62 features. While a more rigorous test could be devised for determining the number of features to use, it would likely not make much difference in overall model performance since the differences in ROC AUC score between different models was only a few thousandths. The model selected from the trial, with the modification of using the 62 features described here, was then saved for use in the Helena Fire site study described in chapter 3.

Figure 2.5 – Shows the model performance, as ROC AUC, resulting from using different numbers of features with the model. For each marker shown, the \( n \) most influential features, as determined by the parameter tuning trial, were used to train and validate the model. Model performance improved rapidly up until about 20 crowns, after which point the score appears to level out. The best score, of about 0.9988, occurs using 62 features (highlighted in red).
2.3.2 Amount of Training Data

In order to determine whether or not the model would benefit from more hand classified data a trial was performed using different fractions of the data set for training. Figure 2.6 shows ROC AUC as a function of crowns used to train the model. It seems there is a leveling out somewhere around 9000 crowns, indicating that there were enough crowns in the training data set.

Figure 2.6 – Shows the model performance, as ROC AUC, resulting from training the model on different sized data sets. Model performance improved rapidly as more crowns were added up until around 2,000 crowns, after which there is a slow improvement until about 9,000 crowns at which point the score appears to level out. The best score, occurs around 9,700 or so.
Table 2.3 – Results using each tile and year combination as a validation holdout. Tile 10TDL0480045075 for the year 2018 does not contain any dead trees. ROC UAC score and Log Loss (as implemented in Scikit-Learn) cannot be computed with fewer than two classes represented in a data set.

<table>
<thead>
<tr>
<th>Name</th>
<th>Year</th>
<th>ROC AUC</th>
<th>Accuracy</th>
<th>Log Loss</th>
</tr>
</thead>
<tbody>
<tr>
<td>10TEL0509245547</td>
<td>2022</td>
<td>1.00000</td>
<td>1.00000</td>
<td>0.026527</td>
</tr>
<tr>
<td>10TDL0458245240</td>
<td>2022</td>
<td>1.00000</td>
<td>1.00000</td>
<td>0.010503</td>
</tr>
<tr>
<td>10TEK0503244655</td>
<td>2022</td>
<td>1.00000</td>
<td>0.998201</td>
<td>0.052565</td>
</tr>
<tr>
<td>10TDL0464245187</td>
<td>2022</td>
<td>1.00000</td>
<td>0.999433</td>
<td>0.010756</td>
</tr>
<tr>
<td>10TDL0480045075</td>
<td>2020</td>
<td>1.00000</td>
<td>1.00000</td>
<td>0.001982</td>
</tr>
<tr>
<td>10TDL0480045075</td>
<td>2018</td>
<td>NaN</td>
<td>1.00000</td>
<td>NaN</td>
</tr>
<tr>
<td>10TEK0500244992</td>
<td>2022</td>
<td>1.00000</td>
<td>0.99571</td>
<td>0.005549</td>
</tr>
<tr>
<td>10TDL0488245360</td>
<td>2022</td>
<td>0.99904</td>
<td>0.985897</td>
<td>0.145979</td>
</tr>
</tbody>
</table>

2.3.3 Spatially Dependent Bias

In order to look for biases introduced by differences in crown populations based on location or other tile-specific traits, the trained model was run on a tile by tile basis. The results (given in table 2.3) were scored using ROC AUC, accuracy, and log-loss. Log-loss was used because I wanted to evaluate how certain the model is in its correct and incorrect class probabilities. Log-loss compares the prediction probabilities to the labels and penalizes them based on distance. For example, a true positive prediction which is more certain (e.g. a prediction probability of 0.9) is penalized less than a true positive prediction that is less certain (e.g. 0.6). A log-loss of zero is a perfect score.

A test was performed using the accuracy scores to determine if there was a statistical difference in the scores of any of the tiles and the expected scores derived by repeatedly drawing random sample populations from the entire population of labeled crowns. Random sample populations (n=131,072), with 512 crowns in each popula-
tion, were drawn from the pool of all labeled crowns from all tiles. Model predictions were made and scored for each sample population. A Bayesian model was then used to test how probable it is for the score of a given tile to be from the distribution underlying the scores of random populations.

Weakly informative priors for mean and standard deviation were used \( \mu \sim N(0.9, 0.01) \) (truncated, \( 0 < a, b < 1 \)) and \( \sigma \sim N(0.3, 0.0025) \). Scores of the sample populations were then used to calculate the likelihood. Finally samples were drawn from the posterior predictive distribution and the 95% highest density interval (HDI) was found.

Figure 2.7 shows the tile scores in comparison to the population of random sample scores. Only one tile falls outside of the 95% credible interval, all other tile were (far) within the 95% credible interval. However, the one tile (10TDL0488245360) that did fall outside of the 95% credible interval of predicted scores, was far outside of the 95% credible interval. It is actually outside of the 99.95% credible interval. As such, the probability that, in tile 10TDL0488245360, the features considered by the model relate to the labels in the same way as in the general population of crowns is less than 0.05%. Put another way, there is a 99.95% probability that there is a non-random cause for the poorer model performance seen on tile 10TDL0488245360.

2.3.4 Quantifying Uncertainty in Prediction Probabilities

Once trained, the model will be used to make predictions on data sets that do not have labels. With no labels, we do not know if the model is right or wrong for a given tree crown. In order to make useful inferences about the mortality rate in unlabeled data it would be helpful to have some way of estimating our uncertainty.

The RF model prediction probabilities reflect the uncertainty the model has surrounding its assignment of class. One would expect that in general, the model should have low certainty about its decision for crowns that it gets wrong. This is mostly
Figure 2.7 – Scores of tiles (green dotted lines) are shown over the histogram (grey) of model prediction scores of random samples of crowns taken from the pooled data set of all labeled crowns. Pink vertical lines show the bounds of the 95% credible interval of the underlying distribution of scores from random samples. Only one tile (10TDL0488245360) falls outside of the 95% credible interval.

the case, as can be seen in Figure 2.8. Given this fact, one means of estimating uncertainty of mortality in a data set would be to produce sample populations representing the probable combinations of living and dead trees. For each population, a crown would be assigned as either alive or dead by drawing from a distribution based on the RF models prediction probability, and some measure of the uncertainty of the value of the prediction probability. There is no obvious quantity to use as this measure of uncertainty of the prediction probability. We can estimate a value by looking at how far from 0.5 (the cutoff point between alive and dead predictions)
the prediction probabilities of correct and incorrect predictions are (as seen in Figure 2.8) and finding a value for $\sigma$ such that if one draws values from $\mathcal{N}(\mu, \sigma^2)$ incorrect classifications will frequently be classified correctly, and correct classifications will continue to be classified correctly the majority of the time.

In order to find a suitable value for $\sigma$, the mean plus the standard distribution of the distance from 0.5 of the incorrect prediction probabilities (shown in the legend of Figure 2.8) was used. This gave a value of $\sigma = 0.30364547029993105$. Using this value, 5 out of 6 wrong predictions (83.33%) are within $\sigma$ of 0.5 but only 34 out of the 2522 correct predictions (1.35%) were within $\sigma$ of 0.5. This should lead to frequent correction of incorrect classifications while seldom affecting the correct classifications.

This value of sigma, and the use of a Gaussian distribution for drawing population samples is not presented as the best method for achieving uncertainty estimates. It is likely that some other value of sigma, or perhaps a different distribution would have better results. Furthermore, it would be ideal to measure the rate of conversion from incorrect to correct, and vice-versa when crown classes are sampled in this way.

In future implementations of the model, the step described in this section will be replaced by calibrating the classifier to make the prediction probabilities more realistic, then using conformal prediction [2] to create statistically rigorous uncertainty intervals.
Figure 2.8 – Kernel densities showing the distance of prediction probabilities from the classification threshold (0.5) grouped by correct and incorrect predictions. Correct predictions (blue line) tend to be very far from the threshold (mean distance = 0.489, where 5 is the maximum possible). Distances of the probabilities of incorrect predictions, on the other hand, are more disperse and, on average, much closer to the threshold (mean = 0.175). Note that negative values and those over 0.5 are artifacts of the kernel density function fitting. Actual values all lie between 0 and 0.5.
2.4 Discussion

The best performing model was a wide, but moderately shallow, Random Forest, that is to say, the model contained a large number of classifiers (111) with a max depth (16) in the middle of the range of values trialed (2–32). This is in line with the parameterization rationale discussed in Section 2.2.6.

2.4.1 Model Selection

After the model had been selected the results of the model were reproducible (so long as the same random seed is used), however, the process of selecting the model was not. Each time the parameters were tuned the outcome was slightly different. It may be that given a very large number of trials the tuning process would eventually converge on an optimum model, but for practical purposes, the results of the tuning process will be unique each time.

In order to assess the variability in models selected by the tuning process, a small tuning experiment was conducted in which the model was tuned ten times. In each tuning run 10,000 trials were conducted. This resulted in eleven tuned models total, ten runs done in the experiment, plus the original tuning run which produced the model used above. Only ten tuning runs were conducted because tuning is a time consuming process. The best trial from each log was recorded.

Though the number of trials is small (n=11), some general trends were suggested by the outcomes. Firstly, in all eleven trials RF models outperformed the SVC models. The parameters for the RF models were mostly well clustered for the best models from each of the eleven trials (see Figure 2.9).

The maximum depth of trees ranged from 11–16 (median value of 13). This is consistent with the findings of Nadi et al. (2019) [30], who ran tests for different
Figure 2.9 – The frequency of occurrence of parameter values for the model tuning runs (n=11). Parameter values are shown on the y-axis. Variation along the x-axis is jitter introduced to make the individual data points distinct. The number of estimators was allowed to range from 2–128. Values were clustered at the high end of this range (rf.n_estimators ranged from 87–128, with a median value of 114). This suggests that if the range had been set higher some models may have ended up with more estimators.

informative features versus the depth of the trees using multiple data sets and found that there tended to be a depth of around 14, after which the accuracy of the RFs did not increase. The number of decision trees was allowed to range from 2–128 in the tuning process. As can be seen in Figure 2.9 the number of estimators was clustered at the top of this range (rf.n_estimators ranged from 87–128, with a median value of 114). It therefore seems likely that if the range had been set higher some models may have ended up with more estimators. Nadi et al. (2019) [30] found that the number of trees needed to reach maximum accuracy goes up with the number of informative features. In this model I have used 62 features—which is not a small number of features.

It is possible that the limitation that was imposed on number of estimators may be detrimental to model performance. The number of estimators is the most important
parameter in the RF model (see Figure 2.10). However, the model scores were so high, any reductions in the performance would be marginal.

It is worth noting that there is strong correlation between some of the most important features used in the model (figure 2.11). As such, some of the features are somewhat redundant. Perhaps the model could be simplified without much decrease in performance by removing some of the features based on correlation. However, the model is not computationally challenging as is and works well, so no effort was made in this regards.

It appears that there existed a number of models, which though unique, were roughly equivalent with regards to performance. The ROC AUC scores of the 11 models produced by the tuning experiment all had a ROC AUC score within 1/1000th of one another (min: 0.9980, max: 0.9988, median: 0.9983). Additionally, in Figure 2.12 it can be seen that, within model tuning runs, many combinations of \texttt{rf.n_estimators} and \texttt{rf.max_depth} fell within areas of very high ROC AUC.
Figure 2.10 – Hyperparameter importances for the first (top) and last (bottom) tuning trials. In the first trial the number of estimators was relatively more important than in the last (0.66 and 0.52) and correspondingly the maximum depth of trees was less important in the first than in the last (0.28 and 0.43). The maximum number of features used per tree was relatively unimportant in both runs (0.06 and 0.05). Relative importances for other trials were not recorded.
Figure 2.11 – Correlation between the 62 features used in the model. Features were sorted by importance, with correlation between the most important features shown in the upper left corner of the triangle.
Figure 2.12 – Contour plots of ROC AUC resulting from different combinations of $rf_{n\text{_estimators}}$ and $rf_{max\text{_depth}}$ for the first (top) and last (bottom) tuning trials. The y-axis corresponds to the number of estimators, and the x-axis to the maximum permitted depth of estimators. ROC-AUC is shown by color, with deeper blue corresponding to higher scores. The pattern of points shows combinations of the two values visited during the tuning process. It can be seen that in both tuning runs, there were many models with extremely high ROC-AUC scores.
2.4.2 Registration error

Registration error is one of largest problems facing the model. Misalignment of trees in imagery with their corresponding crowns can result in misclassifications. NAIP imagery is specified as having horizontal accuracy to within 6 m [31]. Because the study area is quite rugged it is not unreasonable to suspect that the error in this case may be at the outer bounds of what is permitted. Specifications for USGS 3DEP Q1 data require RMSE vertical accuracy of 10 cm or less, I was unable to find a specification for horizontal accuracy, however American Society for Photogrammetry and Remote Sensing (ASPRS) specifications suggest that horizontal accuracy should be within 13 cm [37].

From visual inspection of imagery and tree polygons the LiDAR derived tree polygons and NAIP imagery appear to be within a few meters of one another for the most part, meaning that tree crowns in the imagery were mostly found within the crown polygon. The angle of photos used at a particular location affects how much of the tree falls within the polygon. Trees viewed from directly above tend to line up quite well, while those viewed from a low angle tend to stick out of the polygon more. The photo angle for a given location changes from year to year. The method of feature engineering used should provide some degree of resilience to registration error, as long at the pixels within the tree are predominantly either ground, shadow, or belong to the correct tree.

Here, polygons were labeled by a single person (me) based on subjective perception as to whether a tree was dead or not. Ideally, in lieu of a large set of ground-truthed data (which is in this, and many cases unobtainable) labeling of training and validation polygons should use a consensus more than one person. Also, it is worth noting that the training data tends to omit the most difficult cases, because only crowns which were discernibly alive or dead in the view of the labeler were included.
Generally the model learns to label the data in the way the human/s who provide the training data do. Along with registration error the labeling process is likely where the most improvement could be made to the model.

2.4.3 Geographic Bias

In Section 2.3.3, we saw that the prediction scores of most tiles fell within the range that would be expected given a random sample of tree crowns from the study area.

Tile 10TDL0488245360 was an outlier with regards to model scores, with a worse score that would be expected. Tile 10TDL0488245360, along with 10TEK0503244655 (which was the second worst scoring tile, though it was well within the range of score that would be expected) were the only tiles for which false negatives were more common than false positives. (This can be seen in the confusion matrices presented in appendix B). This suggests that something about these tiles causes some living trees to look more like they are dead to the model. It is also possible that this discrepancy could arise from labeling errors, or at least differences in perception on the part of the labeler.

2.4.4 Epistemology of Dead Trees

An important consideration is what it actually means for a tree to be dead. In labeling tree crowns from remotely sensed images, the working definition of a dead tree used in this study is something like, “a tree of which the crown appears, in the imagery, to be dead, to me”. It is assumed that the dead crown is indicative of a dead stem. Information as to whether or not other less visible parts of the tree are alive is not available. This is not a very good definition, however, it is not markedly different
from other arial tree mortality survey methods (e.g. ADS). Perhaps, *damaged* would be a better term than *dead*, but damaged is extremely vague.

Many of the trees in the study area, particularly hardwoods, can regrow from the root collar even if the stem is dead (e.g. *Quercus spp* [9, 39], as well as *Lithocarpus densiflora* and *Arbutus menziesii* [39]). Conifers on the other hand tend to be dependent on seed for regeneration of sites [39]. It would be useful in future models to differentiate between species (at least between conifers and hardwoods) in order to predict whether regeneration of the same individual, or recruitment of new seedlings is occurring when crowns previously believed to be dead recover.

Growth of shrubs such as *Ceanothus* may confuse the model. Often *Ceanothus* biomass increases rapidly after fire [7]. There is no specific evidence of this occurring here, but logically it seems that it might be a problem.

### 2.5 Conclusions

In this study I set out to provide a model for automated detection of tree mortality which performs roughly as well as a human technician. The results of this chapter show that quantifying tree mortality using an RF model and the combination of LiDAR and arial imagery (in this case 60 cm resolution RGBI NAIP imagery) is a viable method which produces results comparable in accuracy to the labeled data provided for training. It also demonstrates that the RF model is not particularly sensitive to variations in its parameters due to the tuning process. Attention should however be payed to geographic and/or plant community differences between areas used in training and areas for which predictions will be made (see Section 2.4.3). For usage over large areas it may make sense to train several models for different geographic settings and/or plant communities and use a meta-model to assign the correct model to given areas.
The next chapter will provide an example case in which tree mortality is compared across areas with differing recent fire history and landforms.
CHAPTER 3

TREE MORTALITY VARIES BY LANDFORM AND FIRE SEVERITY

3.1 Introduction

This chapter presents a simplistic application of the model described in chapter 2. I wanted to know: 1) Using the model, is it possible to discern differences in tree mortality based on recent fire history? 2) Using the model, is it possible to discern differences in tree mortality based on topographic position? 3) What is the interaction between topographic position and recent fire history in determining mortality rates?

This study does not consider short-term meteorological factors. It does not explicitly account for spatial autocorrelation, however the sampling techniques used likely mitigate the effects. These are important omissions and will be considered in future work. As mentioned in chapter 2, it is difficult to determine, from imagery, if a tree is actually dead or not. For a description of what is meant by mortality please see section 2.4.4 in chapter 2.

3.2 Site

The study site consists of an area affected by two fires in the last decade, the 2017 Helena Fire and the 2021 Monument Fire. The two fires have produced a patchwork of landscapes which have been burned once, twice, or remain unburned.

An area of about 380 km² was selected for the study. The site has been affected by two fires in the last decade, the 2017 Helena Fire and the 2021 Monument Fire.
The area contains about 110 km$^2$ of the Monument Fire footprint, and the entire Helena Fire footprint (about 80 km$^2$). There is about 14 km$^2$ of overlap between the two fires.

The 30 year normal precipitation for the site is 1163 mm (std=259 mm). Vegetation across the site is a fairly typical mix for the Klamath Mountains (KM). Society of American Foresters cover types for the site are listed in table 3.1.

Table 3.1 – Percentage of site landcover by SAF cover type. [38]

<table>
<thead>
<tr>
<th>Vegetation Type</th>
<th>Percentage (%)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Pacific Douglas-fir</td>
<td>37</td>
</tr>
<tr>
<td>Douglas-fir - tanoak - Pacific madrone</td>
<td>17</td>
</tr>
<tr>
<td>Hard chaparral</td>
<td>11</td>
</tr>
<tr>
<td>Canyon live oak</td>
<td>9</td>
</tr>
<tr>
<td>Pacific ponderosa pine - Douglas-fir</td>
<td>8</td>
</tr>
<tr>
<td>Sierra Nevada mixed conifer</td>
<td>5</td>
</tr>
<tr>
<td>Not Forest or Woodland</td>
<td>3</td>
</tr>
<tr>
<td>White fir</td>
<td>2</td>
</tr>
<tr>
<td>California black oak</td>
<td>2</td>
</tr>
<tr>
<td>Oregon white oak</td>
<td>2</td>
</tr>
<tr>
<td>Red fir</td>
<td>2</td>
</tr>
<tr>
<td>Blue oak - digger pine</td>
<td>1</td>
</tr>
<tr>
<td>Pacific ponderosa pine</td>
<td>1</td>
</tr>
<tr>
<td>California mixed subalpine</td>
<td>1</td>
</tr>
</tbody>
</table>

The site was divided into treatment areas based on fire severity (high, medium, low, or unburned) for each of the two fires using RAVG canopy cover loss. This resulted in 16 different treatment combinations, however of the 16 treatment combinations only four, those which contained polygons with areas greater than one ha, were used.

3.3 Methods

Crown polygons were prepared, from the 2019 CarrHirzDeltaFire data set [41], for the entire study area using the lidR package (in R) following the procedure described
Figure 3.1 – Shows site location and treatments. a) Close up of study sites showing areas of the four treatments under consideration. b) Location of site within California. c) Location of site within Trinity County. Note that the Trinity River flows through the southeastern corner of the study site (panel c). Most of the area burned by the Monument fire, but not the Helena fire lies to the south-west of the river.

in section 2.2.3.

Treatments were then ascribed to each crown. Treatment Polygons were created by polygonizing USDA Rapid Assessment of Vegetation Condition After Wildfire (RAVG) burn severity (RAVG) data [44] rasters and keeping polygons larger than 1 ha. This resulted in polygons for only five treatments. One of those treatments, low-unburned, contained only two small polygons (∼1.5 and 2.5 ha) which contained very few crowns. Thus, low-unburned was also dropped resulting in four treatments that were considered; unburned-unburned, unburned-high, high-unburned, and high-high.
After determining the available treatments, crowns were sampled from each treatment for use in the analysis. The number of crowns from each treatment is shown in table 3.2

<table>
<thead>
<tr>
<th>Treatment</th>
<th>Number of Crowns</th>
<th>% of Total</th>
</tr>
</thead>
<tbody>
<tr>
<td>unburned–unburned</td>
<td>14790</td>
<td>31.68%</td>
</tr>
<tr>
<td>unburned–high</td>
<td>14877</td>
<td>31.86%</td>
</tr>
<tr>
<td>high–unburned</td>
<td>14737</td>
<td>31.57%</td>
</tr>
<tr>
<td>high–high</td>
<td>2281</td>
<td>4.89%</td>
</tr>
</tbody>
</table>

A ten-meter buffer was created for each crown polygon. If the buffered crown polygon was found to lie completely within a treatment polygon, it was labeled as belonging to the associated treatment. The rationale behind using the ten-meter buffer was to minimize effects of registration errors between the treatment polygons and crowns. The treatment polygons are derived from RAVG (which is derived from Landsat or Sentinel data [29, 28]). Currently, mean accuracy is listed as within 12 m for both Landsat [20] and Sentinel [8].

I also wanted to consider the effects of landform type on tree mortality. Kane et al. (2015) [23] assigned landforms based on Topographic Position Index (TPI) thresholds for their analysis of burn severity based on topography. I chose to use the geomorphons algorithm [21] rather than TPI. The geomorphons and TPI approaches have been found to be similar in their results [16, 25], and geomorphons is considerably less computationally expensive [21].

A raster containing landforms was calculated using the Whitebox Tools [26] implementation of the geomorphons algorithm, using the CarrHirzDeltaFire 1 m DEM [41] as input. For each crown the geomorphon landform found at its centroid was attached as an attribute. In keeping consistent with the methods of Kane et al.[23] I used outer radii ($r_{outer}$) of 100 m, 250 m, 500 m, 1000 m, and 2000 m and an inner
radius of \( r_{inner} = \frac{r_{outer}}{2} \). (Kane et al. do not specify the inner radius used.) A flatness threshold of 0 was used.

Model input features were generated for each crown, for each year, as described in section 2.2.5. Predictions were made for each crown using the model described in chapter 2. The model predictions were then attached to the crown, treatment, and geomorphon information for analysis.

Simulated populations (\( n=256 \)) were created based on the prediction probabilities from the RF model (\( \hat{p}_{crown_i} \) for \( i = 1 \ldots n_{crows} \)), and the value of \( \sigma \) found in 2.3.4. For each crown a batch of prediction probabilities was drawn from a distribution \( p_{j=1 \ldots 256} \sim \mathcal{N}(\hat{p}_{crown_i}, \sigma^2) \). The probability was then rounded to 1 or 0 (alive or dead) resulting in a \( 256 \times n_{crows} \) matrix where each row contained 256 states (alive or dead) for a particular crown, and each column is a simulated population of living and dead trees.

Crowns were then grouped according to treatment and/or landform. For each group of interest the percentage of dead trees in each column was calculated, resulting in values for percentage dead in each simulated population (\( d_{j=1 \ldots 256} \)). The 95\% HDI was determined for the distribution of percent dead for each group of interest and compared for significance.
3.4 Results

For the site as a whole, including all sampled crowns, the percentage of crowns classified as dead varied substantially by year (figure 3.2, top panel). This may reflect mortality and/or damage due to the Helena and Monument fires (2017, and 2021 respectively). Between 2020 and 2022 there is about a 48% increase in dead crowns. This is the period during which the Monument Fire occurred, which affected about 37% of the crowns in the sample. All of the crowns included in the study which were impacted by the Monument Fire were in areas of high burn severity, as such, it stands to reason that the Monument Fire may have been directly responsible for much of the mortality in this period.

The mortality trend may also be explained by climatic factors as well. The lower panel of figure 3.2 shows annual precipitation for the period from 2006 until 2022. The recovery shown in 2020 (middle vertical red line in lower panel) may also be due to a relatively wet previous year (shown by blue line). This study does not account for climate in a detailed manner, which is a shortcoming that should be addressed in future work.

The sections below will examine the effects of fire severity and landform on mortality.
Figure 3.2 – Top: Boxplot summarizing the distributions of the sample population for the entire site based on the RF model as described in section 3.3. The percentage of crowns classified as dead varies substantially by year. Bottom: Annual rainfall [34] is shown by the blue line. PRISM 30 year normal precipitation [14] for the same point is shown by the dotted red line. The years shown in the upper panel are marked by the vertical red lines.
3.4.1 Mortality by Fire Severity

Figure 3.3 shows the sample distributions for each treatment and year with 95% HDIs. The maximum likelihood estimates for the distributions are given in table 3.3. For each treatment the difference between years is significant, with the exception of 2018 and 2022 in the unburned-unburned treatment which overlap. In a given year the only non-significant difference between treatments is unburned-high and high-high in 2022.

Surprisingly, the unburned-unburned and unburned-high treatments are significantly different in the years 2018 and 2020 (this is especially pronounce in 2018). Since the Monument Fire had not yet occurred, the treatment for both of these groups at this point is the same, unburned. Similarly, high-unburned and high-high show significant difference in these years as well. These differences may be due to differences in composition by landform, aspect, or differences in previous fire history or land treatment. In subsequent sections the first of these possibilities will be explored.

<table>
<thead>
<tr>
<th>Year</th>
<th>All treatments</th>
<th>Unburned</th>
<th>Unburned high</th>
<th>High unburned</th>
<th>High high</th>
</tr>
</thead>
<tbody>
<tr>
<td>2018</td>
<td>48.47</td>
<td>24.48</td>
<td>35.61</td>
<td>81.37</td>
<td>78.10</td>
</tr>
<tr>
<td>2020</td>
<td>40.18</td>
<td>20.81</td>
<td>18.53</td>
<td>77.17</td>
<td>67.28</td>
</tr>
<tr>
<td>2022</td>
<td>59.51</td>
<td>23.33</td>
<td>79.84</td>
<td>71.88</td>
<td>81.68</td>
</tr>
</tbody>
</table>
Figure 3.3 – Shows mortality for each year and treatment under consideration. All differences by year are significant, with the exception of 2018 and 2022 in the unburned — unburned treatment. Maximum likelihood estimates for the distributions shown here can be found in table 3.3.
3.4.2 Mortality by Landform

Across all scales and years depressions and valleys were significantly different from other landforms, and one another. Depressions always had the lowest mortality, and valleys the second lowest.

Ridges were significantly higher than all other land forms in 2018 and 2020 at scale = 2000.

Slopes were significantly different from all other land forms in 2020 at scales of 250 and 500, and in 2022 at scale = 2000, where they were the highest mortality landform.

Summits were significantly different than all other landforms in 2020 at a scale of 2000. Hollows were significantly different from all other landforms in 2022 at scales of 500 and 2000.

Spurs were significantly higher than all other landforms in 2022 at scale = 1000.

Other landforms either formed one or two overlapping clusters depending on year and scale. Members of these clusters were not all necessarily mutually overlapping (though sometimes they are) but in some cases are sequentially overlapping. As can be seen in figure 3.4 the number of groups tends to increase with scale (and, interestingly, also as time progresses). Scale of 2000 consistently has the most separation of groups.

Appendix C shows the complete results of the analysis of crown mortality by landform.
Figure 3.4 – Shows number of significantly different groups by year and geomorphons scale. The number of groups tends to increase with scale (and, interestingly, also as time progresses). Scale of 2000 consistently has the most separation of groups.

3.4.3 Mortality by Severity and Landform

In 2018, only the Helena Fire had occurred, so high-unburned and high-high treatments should be equivalent (the burned groups, as they were burned at high severity), as should unburned-unburned and unburned-high treatments (the unburned groups, as they have not yet burned). As we saw in figure 3.3 however, there is a difference between crown mortality in these two groups.

Figure 3.5 shows the relative composition, by landform, of crowns in each treatment group. Amongst the unburned groups (top row) unburned-unburned has a higher percentage of the more mesic landforms (hollows, valleys, and depressions)
while unburned-high has higher percentage of the more xeric landforms (summit, ridge, spur, slope) across the three scales shown (the exception to this trend is slope, at scale = 100). Amongst the burned groups (bottom row of figure 3.5) The trend is less distinct, but in general the high-unburned treatment group has a higher proportion of the more mesic landforms and high-high have a higher proportion of the more xeric landforms. Ridges and valleys with scale = 2000 do not conform to this trend. The high-high treatment does not contain any crowns found on summit or depression landforms.

In order to determine if the differences seen in the burned and unburned groups in figure 3.3 could be removed by controlling for landform, the simulated populations for the two groups were separated by landform and the 95% HDIs were determined for each landform / treatment combination. The distance between the HDIs of the two groups was measured for each landform and the mean distance between HDIs for the landforms was compared to the distance between HDIs for the group as a whole. (The details and results of this can bee seen in the notebook found in the notebook helena mortality part2 5.ipynb found in the code repository.)

For the unburned group there was a 28% reduction in distance between HDIs after controlling for group. The distance between all landforms however was still significant.

For the burned (before 2019) group (high-unburned and high-high), landform contributes more substantially to the difference between treatments groups. There was a 96% reduction in distance between HDIs after controlling for group. Using a scale of 1000, valleys and slopes did not show a significant difference in percentage of crowns that are dead between the high-unburned and high-high treatments. For ridges, spurs, and hollows, high-unburned was significantly higher than high-high (These results change somewhat with scale, for instance, when this analysis is performed with scale=2000, the difference in hollows becomes insignificant). Summits and depres-
sions that were burned at high intensity in the Helena Fire did not burn again in the Monument fire, thus all summits and depressions that were burned at high intensity at this point belong to the high-unburned treatment. The exact relationships between landforms at different scales and their contributions to the difference between the treatment groups is quite complex, interested readers are encouraged to examine to the notebook helena_mortality_part2_5.ipynb in the code repository for more details.
Figure 3.5 – Relative composition, by landform, of crowns in each treatment group. Amongst the unburned groups (top row) unburned-unburned has a higher percentage of the more mesic landforms (hollows, valleys, and depressions) while unburned-high has higher percentage of the more xeric landforms (summit, ridge, spur, slope) across the three scales shown (the exception to this trend is slope, at scale = 100). Amongst the burned groups (bottom row of figure 3.5) The trend is less distinct, but in general the high-unburned treatment group has a higher proportion of the more mesic landforms and high-high have a higher proportion of the more xeric landforms. Ridges and valleys with scale = 2000 do not conform to this trend. The high-high treatment does not contain any crowns found on summit or depression landforms.
3.5 Conclusions

Due to the lack of crowns that experienced low and medium intensity fire, the effects of fire severity in tree mortality are not particularly informative. The general trend which can be seen in figures 3.2 and 3.3 suggests that after areas were burned at high severity, there was higher tree mortality. This is almost a tautology given that fire severity comes from the RAVG canopy cover loss data set, and the model predicts mortality based on imagery of tree crowns. Recovery is seen in years after high intensity burns in the entire data set (figure 3.2) and across individual treatments figure 3.3. The degree to which this regeneration is recovery of the same trees, or new recruitment of saplings and/or shrubs is a matter in need of more investigation.

More interestingly the study uses geomorphon landforms to assess the impact of topography on crown mortality. Notably, there are significant variations in crown mortality depending on the type of landform. These variations are sensitive to the scale at which the landforms are delineated. In general, more significant differences between groups was found using landforms with large scale parameters. The largest scale parameter used, 2000 m, provided the most separation between groups. For the purposes of creating mortality predictions based on landform this indicates that using a large radius is likely to be more informative.

When all crowns are considered together, unexpected differences arise between treatment groups. In the years 2018 and 2020, the unburned-unburned and unburned-high treatments (which are both unburned before 2021) are significantly different from one another, as are the high-unburned and high-high (both burned once at high severity). Controlling for landform (using scale=1000) about 28% of the difference between the unburned groups was accounted for, and about 96% for the burned groups. This suggests that other attributes not considered here have significant impact on mortality, particularly in unburned areas.
In future works using this model more effort to incorporate climate data should be made. Adding the effects of solar radiation index [24] (SRI) would likely be helpful in explaining some of the unexpected differences between treatment groups. Kane et al. (2015) [23] found that SRI was an important predictor of burn severity. It is an obvious candidate for tree mortality as well. After accounting for some of the unexplained variance a predictive model, based on landform and climate can be constructed.
BIBLIOGRAPHY


Figure A.1 – Society of American Foresters landcover type for 10TEL0509245547.
Figure A.2 – Society of American Foresters landcover type for 10TEK0500244992.
Figure A.3 – Society of American Foresters landcover type for 10TEK0503244655.
Figure A.4 – Society of American Foresters landcover type for 10TDL0480045075.
Figure A.5 – Society of American Foresters landcover type for 10TDL0488245360.
Figure A.6 – Society of American Foresters landcover type for 10TDL0458245240.
Figure A.7 – Society of American Foresters landcover type for 10TDL0464245187.
APPENDIX B

SCORING OF HOLDOUT SITES

Below are confusion matrices resulting from using each tile, year combination as holdout data while training as described in chapter ??.

Table B.1 is the same as 2.3 in chapter ??.

Included here for convenience.
Table B.1 – Results using each tile and year combination as a validation holdout. Tile 10TDL0480045075 for the year 2018 does not contain any dead trees. ROC UAC score and Log Loss (as implemented in Scikit-Learn) cannot be computed with fewer than two classes represented in a data set.

<table>
<thead>
<tr>
<th>Name</th>
<th>Year</th>
<th>ROC AUC</th>
<th>Accuracy</th>
<th>Log Loss</th>
</tr>
</thead>
<tbody>
<tr>
<td>10TEL0509245547</td>
<td>2022</td>
<td>1.000000</td>
<td>1.000000</td>
<td>0.026527</td>
</tr>
<tr>
<td>10TDL0458245240</td>
<td>2022</td>
<td>1.000000</td>
<td>1.000000</td>
<td>0.010503</td>
</tr>
<tr>
<td>10TEK0503244655</td>
<td>2022</td>
<td>1.000000</td>
<td>0.998201</td>
<td>0.052565</td>
</tr>
<tr>
<td>10TDL0464245187</td>
<td>2022</td>
<td>1.000000</td>
<td>0.999433</td>
<td>0.010756</td>
</tr>
<tr>
<td>10TDL0480045075</td>
<td>2020</td>
<td>1.000000</td>
<td>1.000000</td>
<td>0.001982</td>
</tr>
<tr>
<td>10TDL0480045075</td>
<td>2022</td>
<td>1.000000</td>
<td>0.999571</td>
<td>0.005549</td>
</tr>
<tr>
<td>10TDL0480045075</td>
<td>2018</td>
<td>NaN</td>
<td>1.000000</td>
<td>NaN</td>
</tr>
<tr>
<td>10TEK0500244992</td>
<td>2022</td>
<td>1.000000</td>
<td>1.000000</td>
<td>0.003267</td>
</tr>
<tr>
<td>10TDL0488245360</td>
<td>2022</td>
<td>0.999040</td>
<td>0.985897</td>
<td>0.145979</td>
</tr>
</tbody>
</table>
Figure B.1 – Confusion matrix using tile 10TEL0509245547 for 2022 as holdout data. Darker colors indicate a higher count.
Figure B.2 – Confusion matrix using tile 10TDL0458245240 for 2022 as holdout data. Darker colors indicate a higher count.
Figure B.3 – Confusion matrix using tile 10TEK0503244655 for 2022 as holdout data. Darker colors indicate a higher count.
Figure B.4 – Confusion matrix using tile 10TDL0464245187 for 2022 as holdout data. Darker colors indicate a higher count.
Figure B.5 – Confusion matrix using tile 10TDL0480045075 for 2020 as holdout data. Darker colors indicate a higher count.
Figure B.6 – Confusion matrix using tile 10TDL0480045075 for 2022 as holdout data. Darker colors indicate a higher count.
Figure B.7 – Confusion matrix using tile 10TDL0480045075 for 2018 as holdout data. Darker colors indicate a higher count.
Figure B.8 – Confusion matrix using tile 10TEK0500244992 for 2022 as holdout data. Darker colors indicate a higher count.
Figure B.9 – Confusion matrix using tile 10TDL0488245360 for 2022 as holdout data. Darker colors indicate a higher count.
APPENDIX C

MORTALITY BY LANDFORM
Figure C.1 – Percent of crowns dead in simulated populations for 2018 for each landform (using geomorphons with outer radius of 100 and inner radius of 50). 95% HDIs are shown by black bars with labels. Depressions and valleys were significantly different than all other landforms. All other landforms overlapped.
Figure C.2 – Percent of crowns dead in simulated populations for 2018 for each landform (using geomorphons with outer radius of 250 and inner radius of 125). 95% HDIs are shown by black bars with labels. Depressions and valleys were significantly different than all other landforms. All other landforms overlapped some others.
Figure C.3 – Percent of crowns dead in simulated populations for 2018 for each landform (using geomorphons with outer radius of 500 and inner radius of 250). 95% HDIs are shown by black bars with labels. Depressions and valleys were significantly different than all other landforms. Other landform overlapped to various degrees.
Figure C.4 – Percent of crowns dead in simulated populations for 2018 for each landform (using geomorphons with outer radius of 1000 and inner radius of 500). 95% HDIs are shown by black bars with labels. Depressions and valleys were significantly different than all other landforms. Spurs, slopes and hollows formed a cluster, as did summits and ridges.
Figure C.5 – Percent of crowns dead in simulated populations for 2018 for each landform (using geomorphons with outer radius of 2000 and inner radius of 1000). 95% HDIs are shown by black bars with labels. Depressions, valleys, and ridges were significantly different from all other landforms. Summits, spurs, slopes, and hollows formed a cluster.
Figure C.6 – Percent of crowns dead in simulated populations for 2020 for each landform (using geomorphons with outer radius of 100 and inner radius of 50). 95% HDIs are shown by black bars with labels. Depressions and valleys were significantly different than all other landforms. Other landforms overlapped one another.
Figure C.7 – Percent of crowns dead in simulated populations for 2020 for each landform (using geomorphons with outer radius of 250 and inner radius of 125). 95% HDIs are shown by black bars with labels. Depressions and valleys were significantly different than all other landforms. Other landforms overlapped one another to varying degrees.
Figure C.8 – Percent of crowns dead in simulated populations for 2020 for each landform (using geomorphons with outer radius of 500 and inner radius of 250). 95% HDIs are shown by black bars with labels. Depressions, valleys, and slopes were significantly different than all other landforms. Other landforms overlapped one another to varying degrees, though summits were almost significant.
Figure C.9 – Percent of crowns dead in simulated populations for 2020 for each landform (using geomorphons with outer radius of 1000 and inner radius of 500. 95% HDIs are shown by black bars with labels. Depressions and valleys were significantly different than all other landforms. Other landforms overlapped one another to varying degrees.
Figure C.10 – Percent of crowns dead in simulated populations for 2020 for each landform (using geomorphons with outer radius of 2000 and inner radius of 1000). 95% HDIs are shown by black bars with labels. Depressions, valleys, and ridges were significantly different from all other landforms. Summits, spurs, slopes, and hollows formed a cluster.
Figure C.11 – Percent of crowns dead in simulated populations for 2022 for each landform (using geomorphons with outer radius of 100 and inner radius of 50). 95% HDIs are shown by black bars with labels. Depressions and valleys were significantly different than all other landforms. Other landforms overlapped one another to varying degrees.
Figure C.12 – Percent of crowns dead in simulated populations for 2022 for each landform (using geomorphons with outer radius of 250 and inner radius of 125). 95% HDIs are shown by black bars with labels. Depressions and valleys were significantly different than all other landforms. Other landforms overlapped one another to varying degrees.
Figure C.13 – Percent of crowns dead in simulated populations for 2022 for each landform (using geomorphons with outer radius of 500 and inner radius of 250). 95% HDIs are shown by black bars with labels. Depressions, valleys, and hollows were significantly different than all other landforms. Other landforms overlapped one another to varying degrees.
Figure C.14 – Percent of crowns dead in simulated populations for 2022 for each landform (using geomorphons with outer radius of 1000 and inner radius of 500). 95% HDIs are shown by black bars with labels. Depressions, valleys and spurs were significantly different than all other landforms. Other landforms overlapped one another to varying degrees.
Figure C.15 – Percent of crowns dead in simulated populations for 2022 for each landform (using geomorphons with outer radius of 2000 and inner radius of 1000). 95% HDIs are shown by black bars with labels. Depressions, valleys, hollows and slopes were significantly different than all other landforms. Summits, ridges, and spurs formed a mutually overlapping cluster.